

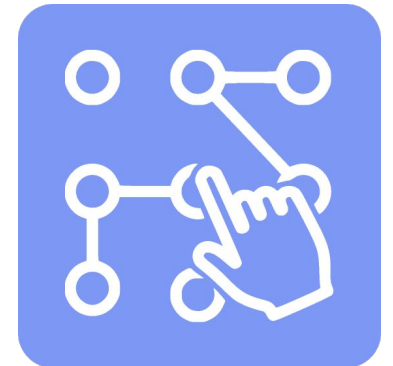
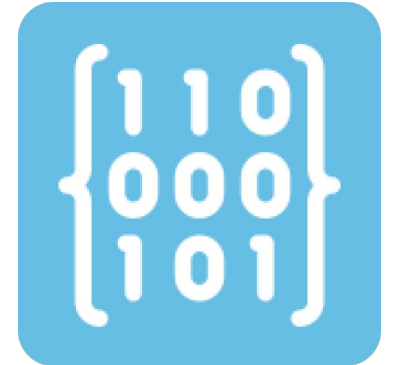
Pandas Practice Session

Relevel
by Unacademy



Questions for Data Transformations using Part 1 :

1. Import Octane.csv Dataset and Calculate the Statistical Properties (mean, median, mode, count, standard deviation, minimum and maximum) or all the numerical columns in the dataframe and store them in separate variables.
 2. Import the dataset from live url :
<https://raw.githubusercontent.com/TrainingByPackt/Data-Science-with-Python/master/Chapter01/Data/student.csv>
-
- A. After importing the dataset from live url, separate out the numerical columns in one dataframe named (df_numeric) and separate out categorical columns in another dataframe named (df_categorical).
 - B. Get the unique values in all the categorical columns and also get the number of unique values in those categorical columns.
 - C. Use value_counts() function on all the categorical columns to get the proportion of values in all the categorical columns. **(value_counts() Example)**
 - D. Use replace function to replace categorical values with numbers for all categorical columns.
 - E. Create a simple numpy array with 3 rows and 3 columns. Assign any numerical values in it and use transform function to apply a transformation to all the values of an array. **(Data Transformation Examples)**



Questions for Data Transformations using Part 2 :

1. Import Data from 'Mall_Customers_BP.csv'.
2. Check the sum of null values across every column.
3. Apply LabelEncoder transformation on the 'Genre' column and store it in a separate data frame.
4. Now, apply OneHotEncoder transformation on the 'Genre' column and store it in a separate data frame.
5. Use `pd.merge()` to merge One Hot Encoded columns with the actual data frame and store it in a separate variable. **(Merge Example)**
6. Create two dummy dictionaries. Convert those dictionaries to dataframes and use `pd.concat()` with different arguments like `join()`, `axis()`, `sort()`, `ignore_index()` and `keys()`.
7. Use the `pd.append()` function on those two dataframes.
8. Use `pd.concat()` to concatenate a pandas data frame and pandas series.
9. Convert the datatypes of some columns using `astype()` function.
10. Use the `downcast()` function to change the data types of all the columns.



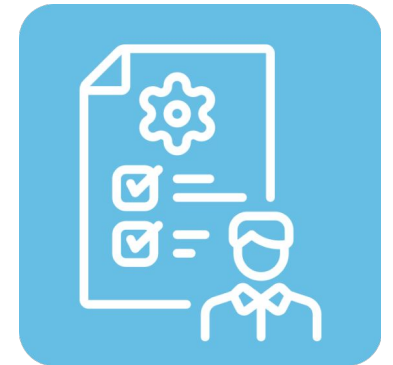
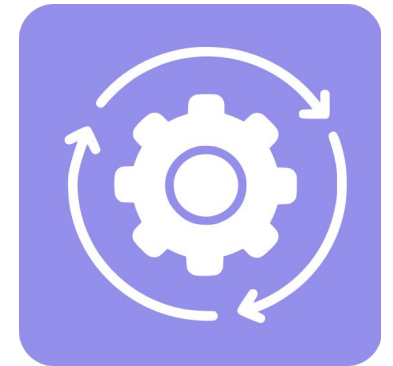
Questions for Data Transformations using Part 3 :

1. Create two dummy dictionaries (available in the solution) and use merge() and join() operations on them. **(Merge and Join Examples)**
2. Import 'nba.csv' dataset. Group the data according to the 'Team' column. **(groupby Example)**
3. Get detailed information about a group using the get_group() function.
4. Group the data according to 'Team' and 'Position' columns simultaneously. **(groupby Example)**
5. Use the Count() function to get the count of entries in every 'Team'.
6. Get the sum, mean, median, maxim, minimum, standard deviation and variance of salary of every member in a team. **(Statistical Functions)**



Questions for Data Transformations using Part 4 :

1. Create a dummy dataframe (Created in solutions). Filter the rows of that dataframe which have values greater than 0.5 in the 'Val' column. **(Filter Condition Examples)**
2. *Apply comparison filter on 'Name' column and explain the concept of comparison filtering on 'Name' column.**(Filter Condition Examples)**
3. Add multiple filtering conditions using OR (|) and AND (&) operators.**(Filter Condition Examples)**
4. Use `isin()` to check for membership of certain values in certain columns.**(Membership Operator Examples)**
5. Explain `startswith()`, `endswith()` and `contains()` function for string columns.
6. Examples of `nlargest()` and `nsmallest()` functions on any numerical column.
7. Use `df.loc[]` and `df.iloc[]` to filter out the dataframes. **(loc and iloc Example (Basics))**
8. Use `sort_values()` to sort the data according to some numerical columns in the dataframe. Use an ascending argument in it.
9. Create a dummy dataframe (created in solutions) and explain `pivot_table()` function on that data.
10. Use `index`, `values` and `aggfunc` argument in `pivot_table()` function.



Questions on requested topics (loc and iloc):

For the following questions you will need an 'Octane.csv' file.

1. Filter out the dataframe rows with Octane number greater than 93.
2. Filter out the dataframe rows with Octane number greater than 93 and Material1 column's value greater than 44.
3. Slice the rows from the data frame with indexes in a particular range.
4. Update the Octane number to 93 where Condition value is greater than 2 and Material 1 is greater than 44.
5. Create a date range from 01/01/2022 to 01/08/2022 with 1 minute frequency.
6. Create a timestamp object for the date 01/11/2022.
7. Create a dummy data frame (available in solutions) and give a timestamp index to it.



Pandas MCQ Questions



Pandas MCQ Questions :

1. A Series object is size mutable

- A. True
- B. False

Answer: B) False

Explanation:

A series object is size mutable. Series Objects are variable in terms of their values, but they are immutable in terms of their sizes. Vector operation refers to the fact that when we apply function or expression to an object, it is applied to each individual item in the object.



Pandas MCQ Questions :

2. A Dataframe object is value mutable.

- A. True
- B. False

Answer: A) True

Explanation:

Sequence Objects are mutable in terms of their values, but they are not mutable in terms of their sizes. When we apply a function or expression to an object, it is applied to each individual item in the object, which is known as vector operation.



Pandas MCQ Questions :

3. Amongst which of the following is / are used to analyze the data in pandas.

- A. Dataframe
- B. Series
- C. Both A and B
- D. None of the mentioned above

Answer: C) Both A and B

Explanation:

We can use series and dataframe to analyze the data in Pandas. Series is one-dimensional labeled array that can store any data type like integers, strings, floating-point numbers, Python objects, etc. A DataFrame is a 2-dimensional labeled data structure with columns that can be of a variety of different kinds. We can think of it as a spreadsheet, a SQL table, or a dict of Series objects. It is one of the most widely used Pandas objects.



Pandas MCQ Questions :

4. During the Execution of following code, what will be the response, we get-

```
import pandas as pd
s =pd.Series([1,2,3,4,5],index= ['a','b','c','d','e'])
print(s['f'])
```

Copy

- A. KeyError
- B. IndexError
- C. ValueError
- D. None of the above mentioned

Answer: C) ValueError

Explanation:

We will get ValueError during the execution of above mentioned code.



Pandas MCQ Questions :

5. Amongst which of the following can be used to create various inputs using pandas DataFrame.
- A. Lists,dict
 - B. Series
 - C. Numpy ndarrays and Another DataFrame
 - D. All of the above mentioned

Answer: D) All of the above mentioned

Explanation:

A pandas DataFrame can be created using various inputs like Lists, dict, Series, Numpy ndarrays, Another DataFrame.



Pandas MCQ Questions :

6. Observe the following code and identify what will be the output when run following code -

```
Import pandas as pd
Import numpy as np

df = pd.DataFrame(np.array([[4,6,9],[5,1,3]]))

print(df.shape)
```

- A. SyntaxError: invalid syntax
- B. KeyError
- C. IndexError
- D. None of the mentioned above

Answer: A) SyntaxError: invalid syntax

Explanation:

When we run the code, invalid syntax error will be reflected.



Thank You